CoSAI
Commission on Sustainable Agriculture Intensification

POLICY BRIEF #3

December 2021

Mining the gaps: Using machine learning to map 1.2 million agri-food publications from the Global South

Photo: C. de Bode/CGIAR

The evidence base on agri-food systems is growing exponentially. A CoSAI commissioned study has applied artificial intelligence to mine more than 1.2 million publications for data, creating a clearer picture of what research has been conducted on small-scale farming and post-production systems from 2000 to the present, and where evidence gaps exist.

Actions needed

- **Research and innovation for agri-food systems should routinely integrate measurements on social equity and health outcomes.** Only a fraction of publications focus on outcomes related to people, such as health and nutrition. The gaps are stark around social equity and inclusion outcomes, such as for women and elderly, indigenous and youth populations.

- **Research and innovation organizations should prioritize programs that go beyond measuring farm and household level outcomes.** There has been relatively little attention to landscape or macro level analyses that are especially important for the natural environment.

- **Research organizations should fast-track research on ecosystems, biodiversity and climate change in various climate zones.** Research on ecosystem services is limited compared to research on technological and socio-economic innovations.

- **Funders should invest in opportunities to increase global research efficiency through identifying and sharing research.** South–South cross-learning increases efficiency and the speed of innovation – and most research on Global South agriculture is being led by researchers in the Global South. Better platforms and toolkits using machine learning will help researchers and decision makers use existing data better.

The challenge: The evidence on agri-food systems is vast and scattered

We are entering a new era in agriculture, one that looks beyond a purely production-oriented vision to a vision of agri-food systems that prioritize people's livelihoods and nutrition as well as environmental and climate outcomes. Agriculture is now a node that touches many issues and disciplines. A bird's-eye view is needed in order to make informed policy and investment decisions.

Earlier work in this area has suggested that the evidence base we have is not fit for the questions we need and want to ask. Additional efforts are needed to help us understand what the current evidence base has found, and where there are gaps.

Keeping up with the evidence is getting harder. Every seven seconds, a new research paper is added to the treasure trove of scientific literature. The volume of research has doubled in the past 10 years. It is increasingly difficult to get an accurate picture of what is out there, especially on a global scale.

Mind the gaps with machine learning models

Advancements in machine learning, a type of artificial intelligence, can help us use the data we already have and keep up with the flood of incoming information. This can be a highly effective way of surfacing relevant insights from a large and representative dataset.

CoSAI's machine learning study looked at the summaries of more than 1.2 million past publications and used these to assess the current landscape of research for the Global South. In order to best assess the immense amount of material available from both English and non-English sources (including development and research organizations, UN agencies, peer-reviewed journals and other publications), the study turned to new technologies that are designed to handle classification tasks with speed and accuracy.

Using Havos.AI machine learning models, the study extracted specific information from each article based on a series of modular questions. The data was then harmonized and cleaned before being presented to human experts for analysis.

Map the gaps with clustering and graphical analysis

Once the data sources are mined, they can be 'mapped' in several ways to better understand the information and how it might be interconnected – especially in the area of agri-food systems, where domains like food and sustainable agriculture tend to overlap.

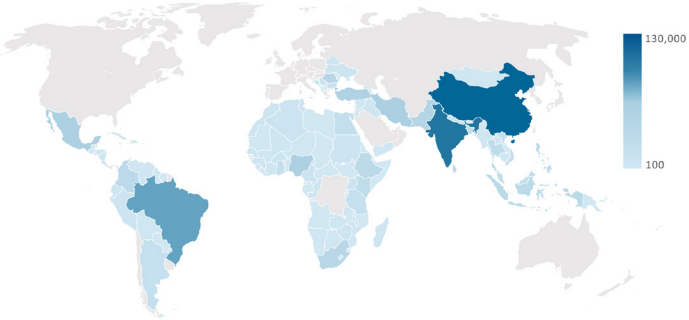
Using a machine learning-assisted clustering technique in which the summary data is examined in a vector space, the CoSAI study applied different algorithms and tested patterns that revealed relationships between different domains. For example, the domains of transportation/ infrastructure and nutrition are connected: thriving markets and the roads that connect them enable the distribution of healthy, safe food that encourages dietary diversity and food security.

The method shows that research tends to cluster together across three pillars of agricultural innovation: **technical**, **socio-economic** and **ecosystem services**. Within each pillar, we identified the top nine intervention areas, based on the quantity of research available in each.

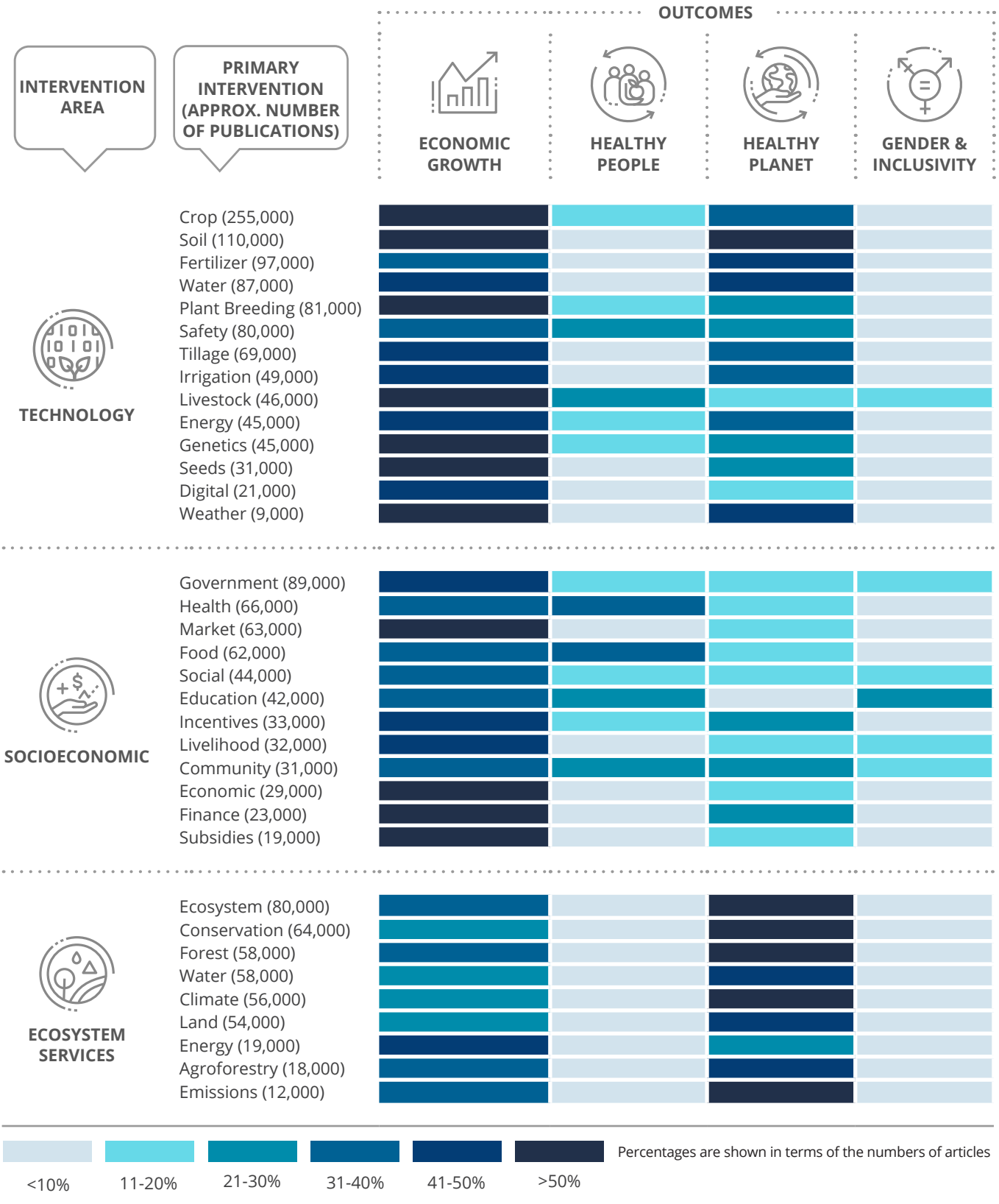
The analysis also produced geographical maps from the extracted data, such as crop research by country and the frequency of articles per region according to different climate zones, as well as demographic breakdowns of study populations.

The analysis highlighted important areas of research that appear to be underfunded. For example, relatively little research is published on **fruits and vegetables** (both in production and post-harvest) – especially in sub-Saharan Africa, and this is a critical area for healthy diets if this does not emerge as a key research priority. The same is true for **biodiversity**, where ecosystem services must play an increasing role in shared natural resources.

A few countries stand out in the mined data on numbers of crop research publications from the Global South



Mapping publications in different intervention areas by the outcomes studied shows gaps in the research evidence – especially in outcomes for people



Close the largest gap: Research on outcomes for people

People-oriented outcomes get little attention in the research literature. For the vast majority of interventions examined in CoSAI's study, less than 10% of research publications focused on health and nutrition outcomes.

An even smaller number of research publications focus on gender and inclusion (social equity) outcomes. Out of 35 intervention areas examined, only six had more than 10% of publications highlighting social equity outcomes, and only one of these was in a 'technical' intervention area (livestock).

Overlapping social factors such as education, socio-economic status, ethnicity, class, caste, age and gender can create interdependent systems of discrimination and disadvantage that reinforce the exclusion of some groups – particularly, but not only, women – from the benefits of agricultural research and innovation. Additional and sustained work in this area will reduce the likelihood of making generalized, homogeneous assumptions for heterogeneous groups, such as small-scale producers.

Conclusions

In new technologies like machine learning, social scientists have incredibly powerful new tools to rapidly mine immense datasets and inform decision making in real time. Graphical maps of this data can contribute additional insight into relationships between and within the data, delivering a more nuanced view of available information, from both a bird's-eye view and zoomed-in perspective.

These tools will be increasingly valuable as policymakers and funders face the need to respond quickly to issues like climate change and food insecurity, which are escalating rapidly. With the rapidly increasing volume of research, machine learning offers a quick way to locate, map and identify critical gaps.

For more information, see the full report at: https://hdl.handle.net/10568/119437

CoSAI
Commission on Sustainable Agriculture Intensification

Supported by:

CGIAR RESEARCH PROGRAM ON Water, Land and Ecosystems

CoSAI is supported by the CGIAR Research Program on Water, Land and Ecosystems and is facilitated by a Secretariat based at the International Water Management Institute headquarters in Colombo, Sri Lanka. WLE is supported by the CGIAR Trust Fund and other donors. CoSAI Commissioners are independent.